DOCUMENT RESUME

ED 443 876		TM 031 521
AUTHOR	Hansen, Joe B.	
TITLE	Student Performance and Student Growth as Measures of Success: An Evaluator's Perspective.	
PUB DATE	2000-04-25	
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).	
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Academic Achievement; Elementary Secondary Education; *Evaluation Methods; *Evaluators; Measurement Techniques; Standards; *Student Evaluation	
IDENTIFIERS	*Cherry Creek School District CO	

ABSTRACT

Measuring the academic achievement of students at the extremes of a grade level distribution accurately poses challenges that may be compounded if a single-level assessment is used for each grade level. Multiple level of achievement measures, such as the Achievement Level Tests (ALT), can help overcome these challenges by increasing the accuracy of the obtained achievement estimate. ALTs provide increased accuracy by providing a measure that centers the score distribution around the student's knowledge level and by providing more items that are at an appropriate level of difficulty for the student. Ceiling and floor effects are less likely, and content validity and curriculum alignment are not at issue because the tests are custom designed to the school district's goal structure. Regression effects are less likely, and ALTs are more growth sensitive than single-level norm referenced tests. The Cherry Creek School District, Colorado, is one of approximately 20 districts in the state and more than 200 in the United States that are using ALTs. ALTS provide a means of addressing the measurement of achievement at the extremes that is necessary to make the new standards-based paradigm effective in improving public education. (SLD)

Student Performance and Student Growth as Measures of Success: An evaluator's perspective

Joe B. Hansen
Teaching Research Division
Western Oregon University
345 N. Monmouth Ave.
Monmouth, OR. 97361
hansenjb@wou.edu

A presentation in the symposium
Measuring Achievement Growth for Students at the Extremes: Educational achievement and progress for high and low performing students.

Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, April 25, 2000

Student Performance and Student Growth as Measures of Success: An evaluator's perspective

Joe B. Hansen
Teaching Research Division
Western Oregon University

Measuring the achievement of students at the extremes of a grade level distribution can be problematic, especially if the assessment devices used are not designed to provide the most accurate measurement for these students. Two types of students or student groups are at risk of being mis-measured: those who score near the lower limit of the test's effective range and those who score at the opposite end of the score range. Students at the lower end are often members of "at-risk" populations, many of whom receive services under Title-1, while those at the upper end may be tagged as "gifted" and receive services for gifted students.

Prior to the widespread implementation of a standards-based approach to education and the accompanying demands for growth for all students, mis-measurement of these groups was of less concern to educators. It was a relatively common practice to focus instruction around the "average student" and move the entire class through the curriculum in a more or less lock-step fashion. This practice had the dual undesirable consequences of failing to provide sufficient challenge to students capable of performing at higher levels, and failing to provide the success experiences needed by lower performing students for them to build upon. Higher ability students were in danger of losing interest in learning the curriculum due to boredom stemming from lack of sufficient challenge, and lower performing students were at risk of becoming frustrated and giving up in despair.

The assessment process most often associated with this older instructional paradigm was norm-referenced, grade level specific, and focused on providing measures of status rather than learning growth. Norm-referenced measures are typically designed to provide the most accurate measurement at the center of their score distribution. This is the level of difficulty where more items are likely to be concentrated, with fewer items addressing the difficulty levels of the lower and upper ends of the distribution. This means that student scores obtained from these outer ranges are subject to higher standard errors. Grade level specific testing compounds this problem by further constraining the range of measurement to that which is deemed appropriate for the grade level.

Standards-based education and the accountability systems that have been put in place to evaluate and report on its effectiveness, require that students at each level of the distribution show improvement over time. At least this is the stated intention of most such systems, even though in practice, these systems sometimes contain flaws that could actually lead educators to focus their attention on students who are at the margin of being proficient, to the exclusion of others within the same proficiency band, but at a lower level.

As an evaluator and research, evaluation and assessment administrator in the public schools for many years, I experienced with each testing cycle, challenges from parents and teachers of high

scoring students, to the "validity" of the scores for their children. Rarely, were such challenges issued by parents or teachers regarding students at the lower end of the score spectrum. Such low scores were more likely to be met with the resignation that these students were simply not learning as well as they should. Teachers would lament: "If only their parents would be more involved...." "If only I could have more time with them." " If only my class size were smaller, perhaps I could help them more." Parents of low achieving children, often from lower socioeconomic strata, and sometimes alienated from the school system themselves, would often accept lower achievement as though it were all they expected. Rarely would parents of low achieving students raise questions with their children's teachers about why those children were not doing well in school.

All too frequently, the economic circumstances of these low achieving students were used as the excuse for their failure to achieve at higher levels. Old studies showing high correlations between income level and achievement test scores have often been used as an excuse for the failure of our education system to help these students achieve their learning potential. Ironically, high achievers are given credit for being great students rather than for coming from affluent and better educated families, where parents emphasize scholastic achievement and set high expectations for their children. The links between student achievement, economic circumstances and the risk of becoming a drop-out have been studied extensively (Goldman, N., Haney, W., and Koffler, S., 1988, Pallas, A.., Natriello, G., and McDill, E., 1989; Levin, H., 1986, Hess, G. A. and Greer, J. [undated]) and will not be examined here. Unfortunately, these studies consistently reveal that minority students tend to come from lower income families and tend to score lower on standardized, norm-referenced achievement tests. In the past this finding has often lead to erroneous, stereotypic conclusions about race and student performance. More recently, evidence is emerging that poverty alone can't explain the gap in achievement between minority and white students. (Viadero, D., 2000) A disturbing finding that warrants much more investigation.

Under the old educational paradigm, where instruction was focused at the middle of the score distribution and norm-referenced measures were used at grade level to evaluate progress, this economics-race-achievement triad was more or less taken for granted and accepted as "the way things are." Factors of socioeconomic status and minority group membership were used as excuses for why many children were not learning at rates commensurate with their age group. Teachers, administrators and entire systems absolved themselves of the responsibility for the learning of these students because of factors that were "beyond their control." Increasingly educators are expected to be accountable for producing higher achievement for all students, regardless of their prior achievement levels, economic circumstances or minority status.

<u>Achievement Growth Across Grades</u>

Experienced evaluators and researchers are familiar with certain ubiquitous growth curves that have been generally accepted as depicting the natural progression of the learning process, especially as it pertains to reading. Figure 1 illustrates the decline in the rate of learning in reading that we often see as students progress from elementary through middle school. The greatest learning seems to be occurring at the earlier grades with little or no additional learning occurring

at the higher grades. In this graph, based on data from a school district in Idaho, the increase in achievement from grade 3 to 4 is greater than the total increase in achievement from grade 4

(Insert figure 1 about here)

through grade 8. Considering that these data are shown in percentile rank units, and that the range of these district level scores is clustered near the middle of the distribution, the jump from third to fourth grade is quite dramatic, while the slow growth from fifth to eighth grade is obvious.

Under the old paradigm, most educators accepted this phenomenon as either reflecting the natural learning and development process, or as depicting the normal way instruction and curriculum were delivered. It was given two common explanations. One, the empty vessel explanation that students have more to learn when they first enter school, therefore it is to be expected that they would show greater growth in the earlier grades and grow less as their brains fill up with knowledge. This explanation assumes that students master basic skills by the time they reach 8th grade. The second, the mis-aligned assessment and curriculum explanation, is that reading is taught directly in the early grades and discontinued after the fifth grade so we shouldn't expect to continue to get gains in reading growth. The latter is as much a condemnation of the curriculum and instructional philosophy of a school or district as it is an explanation; the former, a flawed representation of how learning occurs and how complex basic skills can be.

Analogous to the familiar growth pattern is the phenomenon whereby lower achieving schools, usually with higher levels of poverty and hence more minority students, show greater achievement growth over the course of the year than do higher achieving schools, but fail over time to close the gap between themselves and the higher achieving schools. This gives the principals and staff of the lower achieving schools something to brag about and sometimes gives the district administration a whip to use on the principals of the higher achieving schools to flog them on to higher levels, although this rarely, if ever works and has been known to result in test misuse, abuse and even cheating by teachers and administrators to "get those numbers up."

Some Underlying Issues

There are two sets of issues that arise in these situations I have described. Both need to be considered when evaluating student performance at the extremes. The first set, involves differences in assumptions about learning that occur as a result of the change from a norm-referenced to a standards-based paradigm of education, while the second set relates to the measurement decisions made by educators under these two different paradigms.

Issues related to underlying assumptions about learning

- Old paradigm - assumes that learning is a normally distributed phenomenon and that a normal bell shaped curve of student achievement is inevitable. There will always be students who fail to meet the system's requirements and an equivalent group who exceed the requirements. Most will fall in the middle.

- New paradigm - assumes that all students can and should learn at higher levels than we are accustomed to seeing. While there may be a limit on what any child can learn, we are so far from reaching that limit that we should not see it as a constraint.

Although these two opposing assumptions are important to our understanding of how a standards based approach differs from a more traditional one, my central focus is on issues related to underlying measurement decisions.

Issues related to measurement decisions

The type of assessment used to ascertain whether students are learning well enough may or may not be the most appropriate. The decision about what type of assessment to use is often a reflection of how the educators in charge of assessment view the learning process, and how they plan to use the data. However, the type of assessment chosen, limits the way the data can be interpreted and used. Thus, it is of crucial importance to have both an accurate conception of the learning process and a clear understanding of how assessment data can be used for instructional planning and decision making prior to deciding what type of assessment to use. Each of the following common types of assessment has inherent constraints on its interpretation and use. Some resulting in *relativistic* and others *absolute* judgements about what and how much students are learning.

- Norm-referenced measurement - compares students to the norm group(external) and to each other(internal) as the means of determining whether they have learned enough or not. The result is that all inferences about what the student knows or can do are *relativistic*. Achievement is measured in terms of a student's status relative to the external norm group rather than relative to a standard of performance.

- Criterion referenced measurement - evaluates students' performance in terms of the extent to which they demonstrate mastery on measures of widely accepted learning outcomes. This results in *absolute* judgements about what students know with respect to a given domain defined by the selected outcomes.

- Performance-based assessment - assesses students' performance in terms of how well they can demonstrate knowledge and skills learned through performance of a structured task . Results in *absolute* judgements about what student knows and can do with respect to a *very specific* task, selected to represent a broad domain of learning.

The new paradigm calls for absolute rather than relativistic judgements. The performance standard or benchmark is the criterion against which judgements are made.

Another dimension of the assessment selection process is revealed by the decision to measure either the students' achievement *status* or their *growth*. *Status* refers to a point-in-time measure, showing what the student knows at that time. *Growth*, on the other hand, refers to how much the

student's knowledge or skill level has changed over a specified time period. Status measures can be useful in situations where an absolute level of performance is the overall goal, but are sometimes misused in attempting to demonstrate improved achievement. A typical example is that of a statewide assessment in which the percentage of students at or above proficiency for a particular grade and subject is compared from one year to the next. If a higher percentage of students meets the benchmark the district has shown improvement. An erroneous interpretation would be to say that the district's students had achieved more or grown more that year. Two factors preclude such an interpretation. First is the issue of *comparability*. Two different groups of students are being compared without attention to their similarities or differences on factors that could affect the scores. This year's fourth grade may have a history of scoring higher, or they may have had more training on how to take the state assessments. Second is the *design* of the test itself. It may not be constructed to show growth or may lack the precision to do so. If it is designed to capture the status of students across a broad domain of knowledge, sampling numerous goal areas, with relative few items per goal, and if those items are selected within a relatively narrow range of difficulty, it is not likely to be very sensitive to growth. On the other hand, the test may provide a perfectly adequate measure of whether relatively large groups of students are either above or below the benchmark score for the grade. Nevertheless, we find such measures bing commonly used to make high stakes judgements about individual students relative to their benchmark attainment.

Measuring Growth. Effective evaluation of educational programs and instructional processes requires accurate assessment of the amount of achievement growth attained by students during the instructional period.. This is especially important, since the random selection into programs is rarely feasible and can be considered unethical, depending on circumstances. Therefore, Intact student groups serving as their own controls, often constitute the strongest evaluation design available to us. In order for a test to adequately measure growth in student achievement, it must have certain characteristics. First, the test needs to be constructed so that the difficulty level of the items form a scale that represents growth on the curriculum or learning domain of interest. Item response theory (IRT) provides a means for this type of construction.[1] Second, item selection must be based on the learning goals or standards that form the basis of the curriculum. Third, the assessment should have multiple levels available for the grade level being assessed to optimize the students' opportunities to show growth without the constraints of ceiling and floor effects. In such a scheme, it becomes possible to center the individual student's curriculum knowledge within the range of the test so that the most accurate assessment of knowledge can be obtained. Achievement Level Tests (ALT) afford us a system for this. Fourth, growth related interpretations of the assessment results must be limited to students who have been tested more than once using the same scale within the same content domain. Students who were not present for both testing sessions should not be included in the growth interpretation, although their status at each point can be included in the status measure of their cohort. This type of grouping for

[1] An explanation of Item Response Theory is beyond the scope of this paper. The interested reader is referred to Hambleton, R.K.; Swaminathan, H.; and Rogers, H.J. Fundamentals of Item Response Theory. Sage, 1991, Newbury Park, CA.

reporting purposes is referred to as an "intact" or "matched group," meaning that the same group of students was intact for the period covering both testing sessions and there is a matching post-test score for each pre-test score.

Problems Related to Assessing the Performance of High and Low Achieving Students

Callahan(1992) examined four limitations of standardized tests in evaluating the effectiveness of services for high scoring (gifted) students. These are narrowness of the assessment, invalidity in assessing program goals (content validity/curricular alignment), ceiling effects, and regression to the mean. Callahan indicates that test developers are selective in the curricular goals they choose to represent on their tests, limiting their scope to traditional curricular areas within a narrow range of expectations. "The selection of both the content and the level of thinking required to answer the questions ...reflects the aim of assessing those areas to which most students are exposed in their traditional curricula." (p. 109) This narrowness of the effective range of the test is as likely to result to result in faulty measurement for low achieving as for high achieving students.

Test and Curriculum Alignment. Curricular alignment or content validity can also be a problem when assessing students performing at higher or lower levels, because the students at these levels may not be receiving the same curriculum as their grade level peers. Good teachers will try to work with individual students at a level that is challenging, while offering some opportunity for success - an appropriate instructional level for that student. It's unlikely that a well aligned, single level, content valid test for gifted students would also be well aligned and content valid for students who are struggling with the required curriculum at any given grade.

Ceiling Effects. Ceiling effects occur when there are too few items at the upper end of the effective range of the test to provide an accurate measure of the student's knowledge, therefore artificially suppressing the score when a student gets all or nearly of these items correct. Parents and advocacy groups for gifted students will sometimes criticize a test on the grounds that their students are experiencing ceiling effects. Single level tests are more likely than multi-level tests, to show ceiling effects simply because of the narrowness of the effective range of a single level test. An informal study conducted by members of the Planning, Research and Evaluation Department of Colorado Springs Public Schools revealed that fewer than .5 percent of approximately 2500 eighth grade students were subject to possible ceiling effects on the Achievement level Tests, in spite of protestations from a small group of parents whose children had been identified as gifted and talented.[2]

Floor Effects. Floor effects can be a problem in assessing the performance of low achieving students. The chance level or level at which a student can be expected to score through random guessing, will have a long term mean of about 25 to 30 percent correct on a multiple choice test. Assuming a student is motivated to perform well on the test, we should expect her to score above chance level if she has learned any of the material being tested. Multiple levels allow for low

[2] For further information contact Dr. William Veitch or Dr. Alisabeth Hohn, Colorado Springs School District Eleven.

scoring students to take tests with easier items, thereby increasing the probability of capturing their true knowledge and getting an accurate measure of their achievement.

Regression Effects. Regression to the mean or the regression effect, can also affect the scores of students at the extremes, resulting in exaggerated gains for low achieving students and artificially suppressed gains for high achieving students. Students who score near the lower end of a score distribution may be assumed to have more negative measurement error in their scores than those who score near the middle of the distribution. Conversely, those who score near the upper end of the distribution may be expected to have a disproportionate share of positive measurement error in their scores. On a subsequent testing the redistribution of random error will have the effect of moving the means of each of the groups closer to the mean of the general population from which they are drawn (Talmadge, 1976). When a parent of a high achieving student sees her child's score drop from fall to spring due to regression she may wrongly attribute the decline to faulty instruction, believing the score to be a true measure of what has been learned. Meanwhile, the parent of a low scoring student may believe that his child's score increased due to greater diligence on the part of the student or teacher. It is logical to assume that a test score from a single level test, bearing a larger standard error of measurement, will be more susceptible to false interpretations due to regression than a more accurate score from a multi-level assessment that has a score distribution centered around the student's performance level. This is illustrated in figure 2.

(Insert figure 2 here)

Figure 2 shows the difference between a conventional single grade level test design and an achievement level test design in terms of the increased opportunity for a student to receive a measure that is more nearly centered on his true level of knowledge of the curriculum. Contributing to the more accurate, narrower range of error is the larger number of items available within the difficulty range of the test.

Evaluating Performance and Growth at the Extremes with Level Tests

The ability to effectively analyze the performance of high and low scoring students depends to some extent on the capabilities and options available in the scoring and reporting system. Specifically, it depends on being able to break down the achievement data into subsets by score group and curricular goal area. Exhibit 1 shows a 6th grade Class Report from an actual school. Of course, the name of the school and district have been changed to protect confidentiality. In this report, student scores are arrayed from lowest to highest. Each score is followed by a score range of one standard error. Also shown are four columns corresponding to the four curricular goal areas measured. Percentile scores of 33 or lower are marked "Lo"; between 33 and 66 are "Av" and those above 66 are "Hi." This breakout, enables the teacher to identify more precisely where each individual student's areas of need are for the purpose of targeting instruction. .Exhibit 2 shows a trend report for the same class, which includes the scores for this intact group from the two previous level test administrations. Parent reports of individual student scores, school and district level summaries are also available.

(Insert Exhibits 1 and 2 about here)

My purpose is not to promote Achievement Level Tests, but to point out the type of data that can be most useful in evaluating performance and growth at the extremes. The purpose of having these rich data displays is to evoke appropriate evaluative questioning of the data. For example, one such question might be "is there evidence of a floor or ceiling effect?" To which the answer would be "no." Another useful question would be "are the low scoring students showing similar patterns in terms of their weak areas? " or "are the high scoring students uniformly high on all sub-goals? "

Another useful technique is the use of quartile groupings to see if there are consistent patterns or trends by quartile. Figure 3 shows how tracking quartile groups across grade levels can provide the evaluator useful information, otherwise obscured by grade level means.

(Insert Figure 3 about here)

Figure three shows two kinds of data for an actual school whose name we have changed to conceal its identity. For Brown Caddis Elementary we see an overall pattern of declining growth in mathematics across grades three through eight. However, when the quartile data are plotted against this we see that the pattern varies considerably by quartile grouping. Achievement growth of Q1 students declines sharply from grade 3 through 5 from 10 RIT s to zero. Whereas very little or no growth is occurring for the Q4 students. Q3 students make large gains at grades 3, 5 and 6, but no gains at grade 4. Use of quartile breakouts like these can be very helpful in identifying target areas for improvement.

Summary and Conclusions

Accurately measuring the academic achievement of students at the extremes poses numerous challenges, some of which are compounded if a single level assessment is used for each grade tested. Multiple level achievement measures such as the Achievement Level Tests (ALT) can be helpful in overcoming these challenges by increasing the accuracy of the obtained achievement estimate. These ALTs provide several advantages over single level measures. First, they provide increased accuracy by providing a measure that centers the score distribution around the student's knowledge level and by providing more items that are at an appropriate level of difficulty for the student. Second, ceiling and floor effects are less likely to occur due to the opportunity for a student to receive a more difficult or easier test. Third, content validity or curriculum alignment is not an issue, because the tests are custom designed to the district's goal structure. Fourth, regression effects are less likely because of the improved accuracy which reduces the error component of the observed score. Fifth, because of their improved curricular validity and greater accuracy, and because they employ IRT scaling, ALTs are more growth sensitive than single level NRTs. Cherry Creek School District is one of approximately 20 districts in Colorado and over 200 in the U.S., currently using ALTs to measure and report student growth relative to their curriculum.

Specific, yet simple breakouts of ALT data, showing curricular sub-goal performance for

individuals and groups, accompanied by quartile group trend information can be used to understand more precisely how students at the extremes are performing and how much they are growing.

The challenges posed by measuring achievement at the extremes must be addressed if the new, standards-based paradigm is to be effective in bringing about significant improvements in public education. To the extent that the standards movement is about improving achievement for all students, it must embrace assessment methodologies and reporting strategies that yield high quality, accurate measures of academic achievement for students at all points on the achievement continuum and are capable of measuring and displaying growth in a meaningful way. ALTs provide a means for accomplishing this and should be included as one component in a comprehensive assessment system.

References

Callahan, C.M., 1992. Determining the Effectiveness of Educational Services: assessment issues. In *Challenges in Gifted Education: Developing Potential and Investing in Knowledge for the 21st Century;* EC301 131. ERIC Document Reproduction Service.

Goldman, N., Haney, W., and Koffler, S., 1988. Presentation at AERA Annual Conference, New Orleans.

Hambleton, R.K.; Swaminathan, H.; and Rogers, H.J., 1991. Fundamentals of Item Response Theory. Sage, Newbury Park, CA.

Hess, G. A. and Greer, J. 1988. Categorizing, Sorting and Bending Twigs. Chicago Panel on Public School Policy and Finance. Presentation at AERA Annual Conference, New Orleans.

Pallas, A.., Natriello, G., and McDill, E., 1989; Levin, H., 1986. The Changing Nature of the Disadvantaged Population, Current Dimensions and Future Trends. *Educational Researcher*, June-July, 1989, pp 16-22.

Talmadge, G. K., 1976. The Regression Effect. Technical paper no.3, ESEA Title I Evaluation and Reporting System. RMC Research Corp., USOE, Washington, D.C.

Viadero, D. , 2000. Lags in Minority Achievement Defy Traditional Explanation. *Education Week On the Web*. March 22, 2000.

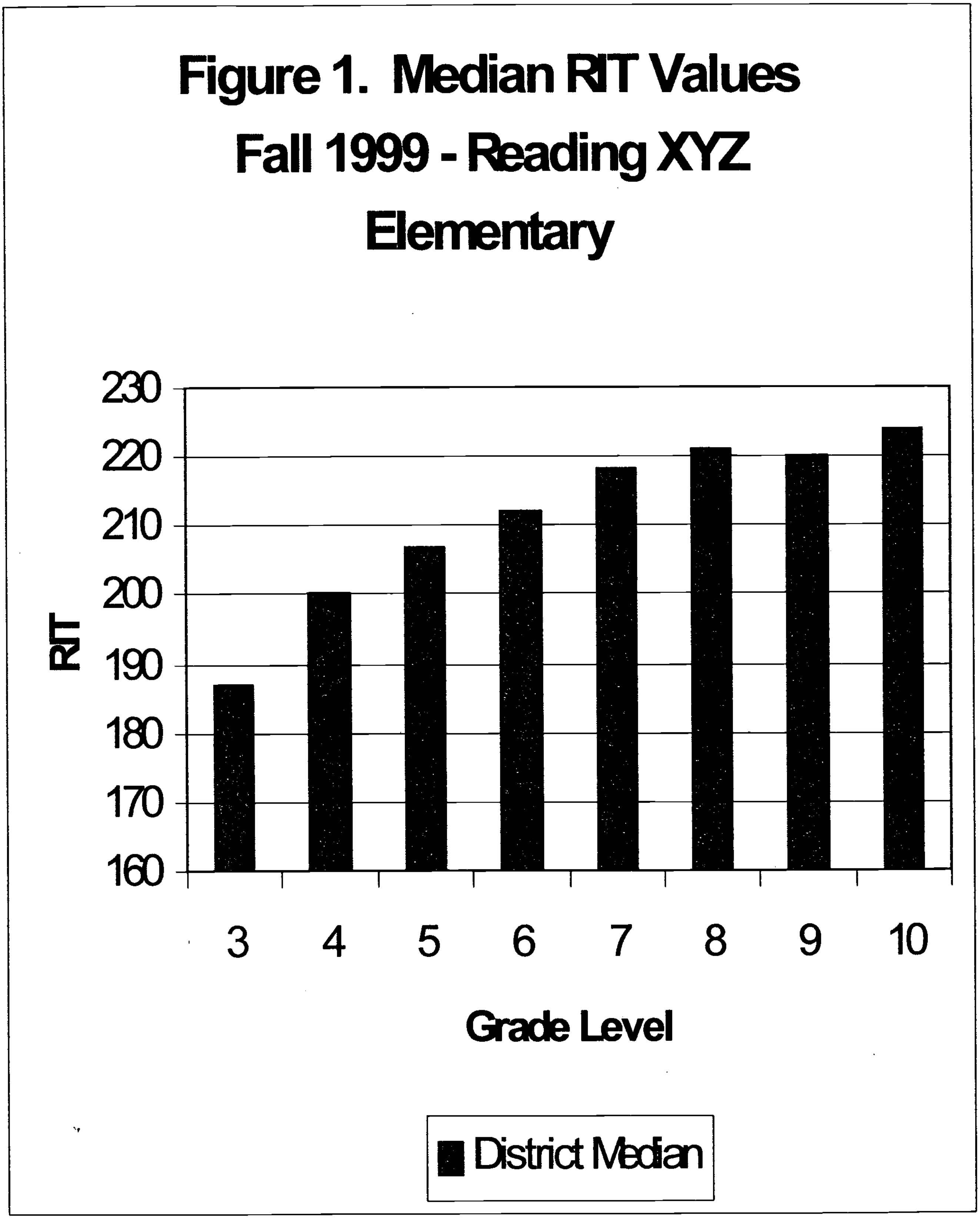


Figure 1. Median RIT Values
Fall 1999 - Reading XYZ
Elementary
230
220
210
200
190
180
170
160
RIT
3
4
5
6
7
8
9
10
Grade Level
District Median

Figure 2. Distribution of Information content in Level Tests Compared with State Assessment

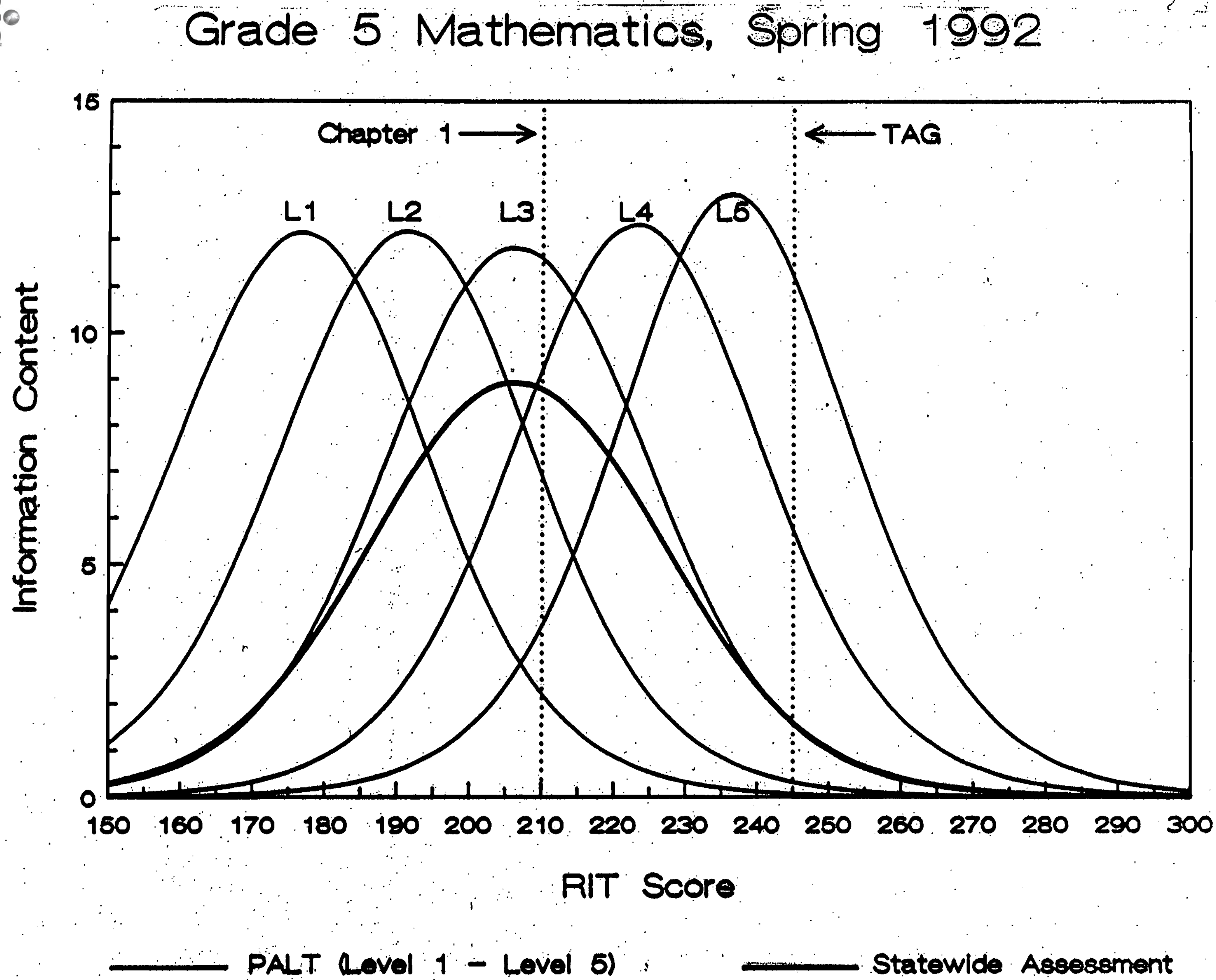


EXHIBIT 1: Level Test Class Report

Achievement Level Test Results

Central School District

Class Report

Spring 1998

Header :80500120

School 400 Washington School

Teacher WARDEN

Reading

ID Number	Name	Grade	Test Form	RIT	Score Range	PCTL	PCTL Range	Goal Performance			
34001592	TREVOR	6	302*4	Special Ed exempt							
34001579	TRAVIS	6	305	201	200-203	16	15-19	Lo	Lo	Lo	Lo
280000027	BRADY	6	364	201	199-202	16	13-18	Lo	Lo	Lo	Lo
34002029	DUSTY	6	364	203	201-205	19	16-22	Lo	Lo	Lo	Lo
24001599	MAGUY	6	305	204	203-206	21	19-24	Lo	Lo	Av	Lo
34001575	ROBERT	6	305	206	204-207	24	21-26	Lo	Lo	Lo	--
34001578	LACEY	6	306	214	212-216	43	38-49	Lo	Av	Av	Av
34001585	LEAH	6	306	215	213-217	46	40-52	Lo	Av	Av	Hi
34001595	COREY	6	307	215	214-217	46	43-52	Av	Av	Av	Av
34001574	RASK	6	306	216	215-218	49	46-55	Av	Av	Av	Av
34001570	JORDAN	6	306	216	215-218	49	46-55	Hi	Av	Av	Lo
34001582	DANIEL	6	307	223	222-225	70	68-76	Av	Hi	Hi	Av
34001596	DANIEL	6	308	224	222-226	73	68-79	Av	Hi	Av	Hi
34001576	RITA	6	307	224	223-226	73	70-79	Hi	Av	Hi	Hi
34001572	KASEY	6	307	224	223-226	73	70-79	Av	Hi	Hi	Hi
280000289	BRADEN	6	308	224	222-226	73	68-79	Hi	Av	Hi	Hi
34001583	COSMO	6	308	226	224-228	79	73-84	Hi	Hi	Hi	Av
280000300	DEX	6	307	226	224-227	79	73-82	Hi	Lo	Hi	Hi
34001577	LAUREN	6	308	228	227-230	84	82-88	Hi	Hi	Hi	Hi
34001584	SHANE	6	308	229	228-231	86	84-90	Hi	Hi	Hi	Hi
34001589	LOUISE	6	307	231	229-232	90	86-91	Hi	Hi	Hi	Hi
34001571	RYAN	6	308	231	229-232	90	86-91	Hi	Hi	Hi	Hi
34001586	WHITNEY	6	307	231	230-233	90	88-93	Hi	Hi	Hi	Hi
34001580	DREW	6	308	232	230-233	91	88-93	Hi	Hi	Hi	Hi
34001591	TRAVIS	6	308	234	232-235	94	91-95	Hi	Hi	Hi	Hi
34001568	TREVOR	6	308	234	232-235	94	91-95	Hi	Av	Hi	Hi
34001593	JESSICA	6	307	234	232-235	94	91-95	Hi	Hi	Hi	Hi
34001581	BAILEY	6	308	234	232-235	94	91-95	Av	Hi	Hi	Hi
34001587	CONNOR	6	308	235	233-237	95	93-96	Hi	Hi	Hi	Hi

	RIT	Goal 1	Goal 2	Goal 3	Goal 4
Number of Students Marked 'Lo'		7	6	4	5
Number of Students Marked 'Av'		6	8	7	5
Number of Students Marked 'Hi'		15	14	17	17
Class Average	221	221	221	224	222
Standard Deviation	10.82	13.73	12.66	12.17	10.89
Class Median	224				

Word Knowl
Literal Comp
Interpretive
Evaluative

Report Date : 9/17/98

Goal Scores:

Hi = Percentile scores > 66
Av = Percentile scores between 66 and 33
Lo = Percentile scores < 33

SC (Special Code) :

2 = ESL Exempt
4 = Sp.Ed. Exempt
5 = Sp. Ed. Modify
6 = Other
7 = Invalid
8 = Out of Level

Note: Since all test scores have some expected error, we suggest the use of score ranges for making your educational decisions. Toward this end, we provide the SCORE RANGE and PCTL RANGE. above:

SCORE RANGE is a range of scores around the student's observed score. If a student were given another level test, his/her score would be within this range most of the time. Based on the test results, students whose score ranges overlap greatly are performing at about the same level.

PCTL RANGE shows the same information as the score range, but in percentile form, for those who are more comfortable using percentile rankings.

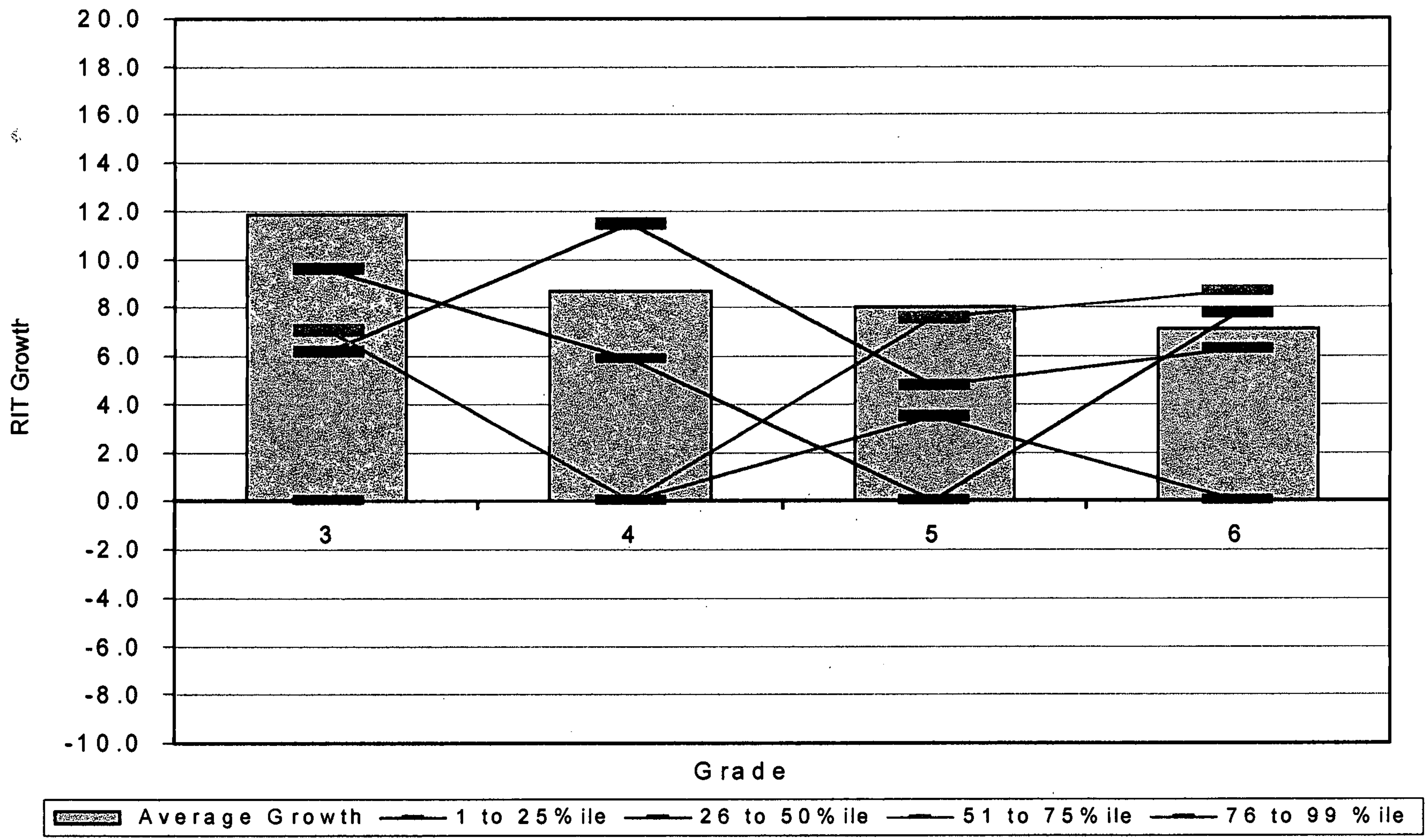
Scoring and Reporting System by NWEA

4/21/2000

Figure 3. Average Growth for Students in each Quartile in Mathematics

Fall 1998 - Spring 1999

Brown Caddis Elementary



17 18

EXHIBIT 2: Level Test Trend Report

Scoring and Reporting System by NWEA

Central School District

Achievement Level Test Results
Trend Report

Report Date: 9/17/98

School # : 400 **School Name:** Washington School
Subject: Mathematics
Header : 80500120
Teacher : WARDEN

			Detailed Test Results														Previous Tests					
			S98						Goal Performance								S97			F97		
Id	Subj	Name	Grd	SC	Msg	Rit	err	%ile	G1	G2	G3	G4	G5	G6	G7	G8	Rit	err	Subj	Rit	err	Subj
24001599	MT	MAGUY	6	0	0	211	3.28	32	A	A	A	H	L				191	4.17	MT	213	3.40	MT
34001574	MT	RASK	6	0	0	224	2.93	64	H	A	A	A	H				221	3.28	MT	222	3.33	MT
34001576	MT	RITA	6	0	0	247	3.75	96	H	H	A	H	H				228	3.06	MT	231	3.23	MT
34001575	MT	ROBERT	6	4	0	162	3.50	1	L	L	L	L	L				196	2.96	MT	193	3.13	MT
34001571	MT	RYAN	6	0	0	235	2.97	86	H	H	H	H	A				237	3.17	MT	237	2.91	MT
34001584	MT	SHANE	6	0	0	237	3.14	88	H	A	H	H	H				220	2.90	MT	222	2.92	MT
34001579	MT	TRAVIS	6	0	0	240	2.91	92	H	H	H	H	H				231	3.22	MT	232	3.03	MT
34001591	MT	TRAVIS	6	0	0	243	2.97	94	H	H	H	H	H				242	3.44	MT	236	2.92	MT
34001568	MT	TREVOR	6	0	0	230	3.22	77	H	H	H	H	H				211	2.99	MT	222	3.33	MT
34001592	MT	TREVOR	6	0	0	214	2.99	39	L	L	A	A	A				209	3.00	MT	210	2.96	MT
34001586	MT	WHITNEY	6	0	0	236	3.10	87	H	H	H	H	H				237	3.74	MT	223	2.93	MT

Term
RIT Avg: 227
RIT Std Dev: 17.57

Note: Scores with a standard error greater than 5 are not shown and are not used in computing the average Rit.

Term 1
RIT Avg: 223
RIT Std Dev: 14.85

Term 2
RIT Avg: 222
RIT Std Dev: 13.25

Header Sheet Sort

Goal Performance:
Hi = Percentile scores > 66
Av = Percentile scores between 66 and 33
Lo = Percentile scores < 33

SC (Special Code):
2 = ESL Exempt
4 = Sp.Ed. Exempt
5 = Sp. Ed. Modify
6 = Other
7 = Invalid
8 = Out of Level

Msg (Message):
1 = school marked invalid
2 = other invalidation
3 = test too easy
4 = test too difficult
5 = too many omits on test
6 = absent
7 = test unscorable
8 = ESL exempt
9 = Special Ed exempt
10 = Special Ed modification
11 = Not used
12 = tested out of grade level
13 = score out of range
14 = test too difficult

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM031521

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: STUDENT PERFORMANCE AND STUDENT GROWTH AS MEASURES OF SUCCESS: AN EVALUATOR'S PERSPECTIVE

Author(s): JOE B. HANSEN

Corporate Source: TEACHING RESEARCH DIVISION, WESTERN OREGON UNIVERSITY

Publication Date: 4/25/00

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: [signature]

Printed Name/Position/Title: JOE B. HANSEN, ASSOC. RESEARCH PROF.

Organization/Address: TEACHING RESEARCH DIV., WESTERN OREGON UNIVERSITY, 345 N. MONMOUTH AV., MONMOUTH, OR 97361

Telephone: 503.838-8820

FAX: 503.838-8150

E-Mail Address: hansenjb@wou.edu

Date: 6/14/00

(over)

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

May 8, 2000

Dear AERA Presenter,

Hopefully, the convention was a productive and rewarding event. As stated in the AERA program, presenters have a responsibility to make their papers readily available. If you haven't done so already, please submit copies of your papers for consideration for inclusion in the ERIC database. We are interested in papers from this year's AERA conference and last year's conference. If you have submitted your paper, you can track its progress at http://ericae.net.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the **2000 and 1999 AERA Conference**. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form enclosed with this letter and send **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can mail your paper to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC is a project of the Department of Measurement, Statistics & Evaluation